JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 22, 523–536 (1968)

# An On-Line Control Scheme Using a Successive Approximation in Policy Space Approach

J. F. BALDWIN AND J. H. SIMS-WILLIAMS

*Department of Theoretical Mechanics, University of Bristol, England*

*Submitted by Richard Bellman*

## 1. INTRODUCTION

1.1 Much theoretical research in recent years has been directed towards the off-line optimization of control systems. The physical realization of the optimal control law, when it can be determined, is nearly always extremely difficult and some form of approximation is necessary. For linear constant coefficient systems with quadratic objective functions, the optimal control law is of a very simple form. If furthermore, steady state conditions are used, the state variables in the feedback are multiplied by constant gains. In this paper a method of on-line optimization is given in which a linear model, quadratic objective function and steady state control are assumed but where the linear model parameters, assumed to remain constant over a small time interval, are updated at successive time intervals. An iterative technique, based upon the method of dynamic programming and successive approximations in policy space, is used to update the feedback gains for use over each time interval. The resulting control, although suboptimal, will often approximate the optimal control within engineering requirements. The goodness of the approximation will depend upon the updating frequency, the rate of change of the linear model parameters and the effectiveness of using steady state control during the final phase of the process.

The scheme is suitable for adaptive systems where the system model assumes a linear form with time varying parameters. The problem of parameter estimation is not dealt with in this paper.

The method relies essentially on two forms of feedback:

(a) The updating of the parameter matrix of the system model at each successive time interval. For non-linear systems the updated parameters will be functions of some state variables of the system.

(b) The dependence of the control on the state of the system over each successive time interval.

Examples are given to illustrate both the iterative technique of finding steady state optimal controls for linear systems with quadratic objective functions and also the application to nonlinear systems.

1.2 We will first give a brief outline of the on-line computing scheme before treating the individual parts in detail.

During the time interval $(t_i, t_{i+1})$ it is assumed that the plant system can be adequately described by the differential equations

$$\frac{d\underline{x}}{dt} = A(t_i)\,\underline{x} + B(t_i)\,\underline{u}; \qquad \underline{x}(t_i) = \underline{x}_i\,, \tag{1}$$

where $\underline{x}$ is an $n$ state vector, $\underline{u}$ an $m$ control vector, $A$ an $(n \times n)$ constant matrix and $B$ an $(n \times m)$ constant matrix. At the time instant $t_i$, the future behavior of the plant is viewed with complete optimism as it is supposed that the matrices $A(t_i)$ and $B(t_i)$ remain constant. The control for the time period $(t_i, t_{i+1})$ is calculated using the iterative scheme

$$\underline{u} = \underset{r\to\infty}{Lt}\ \underline{u}_r \tag{2}$$

$$\underline{u}_{r+1} = K_{r+1}\underline{x} \tag{3}$$

where

$$K_{r+1} = -\frac{B^T(t_i)\,P_r}{2\lambda} \tag{4}$$

$P_r(A(t_i) + B(t_i)\,K_r) + [P_r(A(t_i) + B(t_i)\,K_r)]^T = -2(Q + \lambda K_r K_r^T)$. $K_r$ is an $(n \times n)$ matrix and $K_0$ is given. The scheme is based upon minimizing the functional

$$J = \int_0^\infty (\underline{x}^T Q \underline{x} + \lambda \underline{u}^T \underline{u})\,dt. \tag{6}$$

The iterative scheme possesses approximate quadratic convergence as the optimal solution is approached and will always converge provided $K_0$, the initial guess of the control gain matrix, is such that the system

$$\frac{d\underline{x}}{dt} = (A(t_i) + B(t_i)\,K_0)\,\underline{x} \tag{7}$$

is asymptotically stable with respect to the origin $\underline{o}$. The calculation of this control gain matrix must be performed prior to time $t_i$ for on-line computing. At time $t_{i+1}$ the matrices $A$ and $B$ are updated and a new control gain matrix found using the above iterative scheme with $A(t_i)$ and $B(t_i)$ replaced by $A(t_{i+1})$ and $B(t_{i+1})$.

## 2. Successive Approximations in Policy Space

**2.1** This has been discussed by R. E. Bellman in Refs. [1]–[3]. Consider a system described by the set of differential equations:

$$\frac{d\underline{x}}{dt} = f(\underline{x}, u); \qquad \underline{x}(t_0) = \underline{x}_0 , \tag{8}$$

where $\underline{x}$ is the state vector and $u$ a control variable to be determined so as to minimize the objective function $J$:

$$J = \int_0^\infty \phi(\underline{x}, u)\, dt. \tag{9}$$

By using the method of dynamic programming the solution is found by solving the recurrence relation

$$S(\underline{x}) = \min_{u(\underline{x})}[\phi(\underline{x}, u)\, h + S\{\underline{x} + f(\underline{x}, u)\, h\}, \tag{10}$$

where $h$ is a sufficiently small time interval and

$$S(\underline{x}) = \min_{u(\underline{x})} \int_t^\infty \phi(\underline{x}, u)\, dt \tag{11}$$

with the process starting in state $\underline{x}(t)$ at time $t$.

The approximation in policy space is then given by the iterative scheme

$$\begin{aligned} S_{(r)}(\underline{x}) &= \phi(\underline{x}, u_r)\, h + S_{(r)}\{\underline{x} + f(\underline{x}, u_r)\, h\} \\ u_{r+1} \text{ from: } &\min_{u(\underline{x})}[\phi(\underline{x}, u)\, h + S_{(r)}\{\underline{x} + f(\underline{x}, u)\, h\}] \end{aligned} \tag{12}$$

with $u_0(\underline{x})$ given so that we obtain

$$S_{(r)}(\underline{x}) = \sum_{i=0} \phi\{\underline{X}_i u_r(\underline{X}_i)\}\, h,$$

where

$$\underline{X}_i = \underline{X}_{i-1} + f\{\underline{X}_{i-1}, u_r(\underline{X}_{i-1})\}\, h \tag{13}$$

with $\underline{X}_0 = \underline{x}$; and $u_{r+1}$ calculated using

$$\min_{u(\underline{x})} \left[\phi(\underline{x}, u) + \sum_{i=0} \phi\{\underline{Y}_i, u(\underline{Y}_i)\}\, h\right],$$

where

$$\underline{Y}_i = \underline{Y}_{i-1} + f\{\underline{Y}_{i-1}, u_r(\underline{Y}_{i-1})\}\, h \tag{14}$$

with $\underline{Y}_0 = \underline{x} + f(\underline{x}, u(\underline{x}))\, h$.

This defines a computational algorithm:

(1) Start with given $u_0(\underset{\sim}{x})$.

(2) Determine $u_1(\underset{\sim}{x})$ from (14) and store.

(3) Determine $u_2(\underset{\sim}{x})$ from (14) and store, etc.

The disadvantage of such an algorithm is that, since $u_r$ is a function of the state of the system, a multivariable function must be stored at each stage of the process. This is the familiar problem of the "curse of dimensionality." The above type of scheme can be made the basis of an on-line computing method of control in which storage difficulties are traded for computation and the feasibility of the method depends on the speed of computation. In the method developed below, these problems (storage or speed of computation) are overcome at the expense, in the nonlinear case, of obtaining only a suboptimal feedback control. This suboptimal control can be improved using an on-line computing scheme such as that mentioned briefly above and which will be the subject of a future paper.

**2.2** We now consider the special case of a time invariant linear system with quadratic objective function:

$$\frac{d\underset{\sim}{x}}{dt} = A\underset{\sim}{x} + \underset{\sim}{b}u; \qquad \underset{\sim}{x}(0) = \underset{\sim}{x}_0 \tag{15}$$

$$\min_u J = \int_0^\infty (\underset{\sim}{x}^T Q \underset{\sim}{x} + \lambda u^2)\, dt, \tag{16}$$

where $\lambda$ is a weighting factor and $Q$ a symmetric weighting matrix.

The results will be generalized to systems in which the control is a vector.

From the general treatment given above, the $(r + 1)$th approximation to the optimum control variable $u$ is found using the iterative scheme

$$\min_u \left[(\underset{\sim}{x}^T Q \underset{\sim}{x} + \lambda u^2) + \sum_{i=0} \{\underset{\sim}{Y}_i^T Q \underset{\sim}{Y}_i + \lambda u_r^2(\underset{\sim}{Y}_i)\}\right] h,$$

where

$$\underset{\sim}{Y}_i = \underset{\sim}{Y}_{i-1} + [A\underset{\sim}{Y}_{i-1} + \underset{\sim}{b}u_r(\underset{\sim}{Y}_{i-1})]\, h \tag{17}$$

with $\underset{\sim}{Y}_0 = \underset{\sim}{x} + [A\underset{\sim}{x} + \underset{\sim}{b}u(\underset{\sim}{x})]\, h$. Assume $u_r(\underset{\sim}{x})$ is of the form

$$u_r(\underset{\sim}{x}) = \underset{\sim}{k}_r^T \underset{\sim}{x},$$

where $\underset{\sim}{k}_r$ is an $n$ constant column vector, so that $u_{r+1}$ is found using

$$\min_u \left[(\underset{\sim}{x}^T Q \underset{\sim}{x} + \lambda u^2) + \sum \underset{\sim}{Y}_i^T \{Q + \lambda \underset{\sim}{k}_r \underset{\sim}{k}_r^T\}\, \underset{\sim}{Y}_i\right],$$

where

$$\underline{Y}_i = [I + (A + \underline{b}\underline{k}_r^T)\, h]\, \underline{Y}_{i-1} \tag{18}$$

and

$$\underline{Y}_0 = \underline{x} + [A\underline{x} + \underline{b}u(\underline{x})]\, h$$

Therefore

$$\underline{Y}_i = [I + (A + \underline{b}\underline{k}_r^T)\, h]^i \{\underline{x} + [A\underline{x} + \underline{b}u(\underline{x})]\, h\}. \tag{19}$$

Let

$$C_r = I + (A + \underline{b}\underline{k}_r^T)\, h$$

$$E_r = Q + \lambda \underline{k}_r \underline{k}_r^T.$$

The $u_{r+1}$ satisfying the above minimum condition is thus given as a solution of the equation

$$2\lambda u_{r+1} + \frac{\partial}{\partial u_{r+1}} \left[ \sum_{i=0} \underline{Y}_i^T E_r \underline{Y}_i \right] = 0, \tag{20}$$

giving

$$2\lambda u_{r+1} + h \left\{ \sum_{i=0} \underline{b}^T (C_r^i)^T (E_r + E_r^T)\, C_r^i \right\} \{\underline{x} + (A\underline{x} + \underline{b}u_{r+1})\, h\} = 0. \tag{21}$$

Let

$$h \left\{ \sum_{i=0} \underline{b}^T (C_r^i)^T (E_r + E_r^T)\, C_r^i \right\} = \underline{b}^T P_r\,, \tag{22}$$

then, assuming that $\underline{k}_0$ has been chosen to guarantee convergence of the above summation, $P_r$ satisfies the system of equations

$$P_r - [I + h(A + \underline{b}\underline{k}_r^T)^T]\, P_r [I + h(A + \underline{b}\underline{k}_r^T)] = h(E_r + E_r^T) \tag{23}$$

so that in the limit as $h \to 0$

$$P_r(A + \underline{b}\underline{k}_r^T) + (A + \underline{b}\underline{k}_r^T)^T P_r = -2E_r \tag{24}$$

and

$$S_{(r)} = \int_t^{\infty} (\underline{x}^T Q \underline{x} + \lambda u_r^2)\, dt. \tag{25}$$

This is a system of equations which is solved numerically for $P_r$.

We also obtain from equations (21) and (22)

$$u_{r+1}(\underline{x}) = \frac{-\underline{b}^T P_r (I + Ah)}{2\lambda + h\underline{b}^T P_r \underline{b}}\, \underline{x}. \tag{26}$$

We therefore obtain in the limit as $h \to 0$ that

$$\left.\begin{aligned} u_{r+1} &= \underline{k}_{r+1}^T \underline{x} \\ \text{where} \qquad \underline{k}_{r+1}^T &= -\frac{\underline{b}^T P_r}{2\lambda} \\ P_r(A + \underline{b}\underline{k}_r{}^T) + [P_r(A + \underline{b}\underline{k}_r{}^T)]^T &= -2E_r \end{aligned}\right\} \tag{27}$$

Equations (27) define an iterative procedure for the calculation of $u(\underline{x})$. It is obvious that $P_r$ is a symmetric matrix and therefore, at most, $\frac{1}{2}n(n + 1)$ linear equations must be solved at each stage of the iteration where $n$ is the dimension of the state vector. The convergence of this process will be discussed in a later section.

The iteration scheme is easily generalized for the more general system

$$\frac{d\underline{x}}{dt} = A\underline{x} + B\underline{u} \tag{28}$$

$$\min \int_0^\infty (\underline{x}^T Q \underline{x} + \lambda \underline{u}^T \underline{u})\, dt, \tag{29}$$

where $\underline{u}$ is an $m$-dimensional control vector.

In this case the following formulas are applicable

$$\left.\begin{aligned} \underline{u}_{r+1} &= K_{r+1}\underline{x} \\ \text{where} \quad K_{r+1} &= -\frac{B^T P_r}{2\lambda} \\ \text{and} \quad P_r(A + BK_r) + [P_r(A + BK_r)]^T &= -2(Q + \lambda K_r{}^T K_r). \end{aligned}\right\} \tag{30}$$

The value of the objective function (cost of the process), starting the process in state $\underline{x}$, when the control $\underline{u}_r$ is used, is given by

$$S_{(r)}(\underline{x}) = h \sum \underline{X}_i{}^T \{Q + \lambda K_r{}^T K_r\} \underline{X}_i$$

where

$$\underline{X}_i = \underline{X}_{i-1} + (A\underline{X}_{i-1} + BK_r\underline{X}_{i-1})\, h; \qquad \underline{X}_0 = \underline{x}. \tag{31}$$

Therefore

$$\begin{aligned} S_{(r)}(\underline{x}) &= h \sum \underline{x}^T (C_r{}^i)^T E_r C_r{}^i \underline{x} \\ &= \tfrac{1}{2}\underline{x}^T P_r \underline{x} \end{aligned} \tag{32}$$

using Eqs. (22) and (32) where the modified definitions

$$C_r = I + (A + BK_r)\, h$$

$$E_r = Q + \lambda K_r{}^T K_r$$

are used.

## 3. Connection with the Classical Solution and Convergence

**3.1** The solution of the linear control problem with quadratic cost function

$$J = \int_0^T (\underline{x}^T Q \underline{x} + \lambda \underline{u}^T \underline{u})\, dt \tag{33}$$

is given by

$$\underline{u} = -\frac{1}{\lambda} B^T W \underline{x}, \tag{34}$$

where $W$ is a symmetric matrix whose elements are a function of "time to go" satisfying the matrix Riccati differential equations

$$-\frac{dW}{dt} = WA + A^T W - \frac{1}{\lambda} WBB^T W + Q; \qquad W(T) = [0]. \tag{35}$$

For $T \to \infty$ the steady state solution is given by

$$\underline{u} = -\frac{1}{\lambda} BW_s \underline{x}, \tag{36}$$

where $W_s$ satisfies the quadratic equation

$$W_s A + A^T W_s - \frac{1}{\lambda} W_s BB^T W_s + Q = [0]. \tag{37}$$

Assume $\mathrm{Lt}_{r\to\infty} P_r = P$ to exist and let $K = -(B^T P/2\lambda)$ and $E = Q + \lambda K^T K$, then

$$P(A + BK) + (A + BK)^T P = -2E, \tag{38}$$

and substituting for $K$ gives

$$PA + A^T P - \frac{1}{2\lambda} PBB^T P + 2Q = [0]. \tag{39}$$

Hence

$$P = 2W_s. \tag{40}$$

**3.2** We will now show that, if $P_0$ exists, the sequence $\{S_{(r)}\}$ is monotonic decreasing.

Since $P_0$ exists, $S_{(0)} = \frac{1}{2}\underline{x}^T P_0 \underline{x}$ is finite. From equation (12)

$$\begin{aligned} S_{(r+1)}(\underline{x}) &= (\underline{x}^T Q \underline{x} + \lambda u_{r+1}^2)\, h + S_{(r+1)}\{\underline{x} + (A\underline{x} + \underline{b} u_{r+1})\, h\} \\ &= \tfrac{1}{2}\underline{x}^T P_{r+1} \underline{x} \quad \text{in the limit as} \quad h \to 0. \end{aligned} \tag{41}$$

The method of determining $u_{r+1}(\underline{x})$ ensures that

$$(\underline{x}^T Q \underline{x} + \lambda u_r^2)\, h + S_{(r)}\{\underline{x} + (A\underline{x} + \underline{b}u_r)\, h\}$$
$$\geqslant (\underline{x}^T Q \underline{x} + \lambda u_{r+1}^2)\, h + S_{(r)}\{\underline{x} + (A\underline{x} + \underline{b}u_{r+1})\, h\}. \tag{42}$$

Therefore $\underline{x}^T P_r \underline{x} \geqslant \underline{x}^T P_{r+1} \underline{x}$ by Eqs. (31) and (32). Hence

$$\cdots S_{(r+1)} \leqslant S_{(r)} \leqslant \cdots \leqslant S_{(1)} \leqslant S_{(0)} .$$

The iterative scheme is, therefore, monotonic decreasing provided the initial control $\underline{k}_0{}^T \underline{x}$ is such that the value of the objective function $S_{(0)}$ is finite. $S_{(0)}$ will be finite provided the initial system

$$\frac{d\underline{x}}{dt} = (A + \underline{b}\underline{k}_0{}^T)\, \underline{x}$$

is asymptotically stable with respect to the origin. This follows directly from the exponential form of the solution of a constant coefficient linear system and the asympototic stability property.

The above iteration scheme is equivalent to a Newton-Raphson type of approximation for solving Eq. (37) and therefore the method will exhibit quadratic convergence in the region of the true solution.

## 4. Application to Nonlinear Systems

The application of the above control scheme to nonlinear systems defines two control policies. In the first policy the control is to be regarded as suboptimal and only in certain instances will approximate within reasonable degree of accuracy to the optimal control. A more useful interpretation is that, for nonlinear systems, an on-line control is to be used in which the nonlinear structure of the system is used only to update a linear model. In engineering terms this can be viewed as model feedback which is to be used in conjunction with state feedback. With this interpretation the problem of determining the optimal control at any time instant is localized and can be determined by the iterative scheme developed above. The practicability of this scheme for a particular on-line system must be investigated and may prove to be inadequate, in the same way that a state feedback scheme, based upon a deterministic model will not always yield a feasible control policy. For such cases, more information of the effect of the nonlinear system structure must be gained. Several methods can be used for this purpose, but the advantage of the above scheme is that for many engineering systems it will yield an adequate control which is feasible as an on-line method of computation.

The above iterative technique can also be valuable in determining a first approximation for use with the various gradient techniques for finding optimal control of nonlinear systems.

## 5. Examples

**5.1** *A time invariant linear system.* An example of the type discussed in 2.2 with

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5 & -4 & -3 & -2 \end{bmatrix}; \qquad \underset{\sim}{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\underset{\sim}{Q} = I \qquad \text{and} \qquad \lambda = 0.1 \quad \text{was taken.}$$

Starting from widely different but stable initial control gain vectors, $\underset{\sim}{k} = [4, 0, -3, -2]$ and $\underset{\sim}{k} = [-11, -28, -21, -6]$, the iterative scheme given in Eqs. (27) converged (see Table 5.1.1) in six and five iterations, respectively, to the optimal steady state control gain vector

$$\underset{\sim}{k} = [-0.91608, -7.3673, -8.2588, -3.5243].$$

TABLE 5.1.1

Example of Convergence from Stable Starting Control Gain Vectors $k$

| Iteration number | $k_1$ | $k_2$ | $k_3$ | $k_4$ |
|---|---|---|---|---|
| 0 | 4 | 0 | −3 | −2 |
| 1 | −13.00000 | −29.906250 | −24.625000 | −7.9062500 |
| 2 | −4.9722221 | −15.501044 | −14.230986 | −5.0963160 |
| 3 | −1.7409857 | −9.3046373 | −9.7340069 | −3.9062840 |
| 4 | −.96655239 | −7.5355660 | −8.3976616 | −3.5601385 |
| 5 | −.91629330 | −7.3688825 | −8.2602196 | −3.5246447 |
| 6 | −.91607973 | −7.3673022 | −8.2587696 | −3.5242682 |
| 0 | −11 | −28 | −21 | −6 |
| 1 | −4.0937500 | −14.866211 | −13.216797 | −4.5270996 |
| 2 | −1.4712736 | −9.1181693 | −9.5417301 | −3.7978654 |
| 3 | −.93989585 | −7.5124243 | −8.3800081 | −3.5516345 |
| 4 | −.91612749 | −7.3686582 | −8.2600198 | −3.5245609 |
| 5 | −.91607981 | −7.3673023 | −8.2587697 | −3.5242682 |

TABLE 5.1.2

EXAMPLES OF CONVERGENCE FROM UNSTABLE STARTING CONTROL GAIN VECTORS $k$

| Iteration number | $k_1$ | $k_2$ | $k_3$ | $k_4$ |
|---|---|---|---|---|
| 0 | 0 | 4 | 0 | 0 |
| 1 | −1.0000000 | 6.8000000 | 3.2000000 | −0.90000000 |
| 2 | −.91666667 | 4.3218541 | 3.1725792 | −.76980028 |
| 3 | −.91607981 | 4.2162499 | 2.6589795 | −.95216855 |
| 4 | −.91607978 | 4.1785636 | 2.6593636 | −.94640563 |
| 5 | −.91607978 | 4.1784742 | 2.6592462 | −.94643982 |
| 0 | −5 | −5 | −5 | −5 |
| 1 | −1.7500000 | 18.402986 | 17.212687 | −0.04104470 |
| 2 | −.96759256 | 6.3907959 | 14.616076 | 4.7109372 |
| 3 | −.91630216 | −4.4597917 | 6.6328520 | 3.4909007 |
| 4 | −.91607978 | 3.4445600 | 7.9162361 | 2.1308983 |
| 5 | −.91607978 | 63.928784 | 8.3145160 | −7.9768225 |
| 6 | −.91608003 | 33.401971 | 8.3591600 | −2.8521795 |
| 7 | −.91607978 | 17.557248 | 8.3461313 | −.14866157 |
| 8 | −.91607978 | 8.3249334 | 8.2847985 | 1.5236222 |
| 9 | −.91607978 | −.92970194 | 7.8466481 | 3.5390729 |
| 10 | −.91607978 | 1.1078263 | 2.3058623 | 5.8195139 |
| 11 | −.91607978 | 3.1488863 | 2.9528008 | 4.9693680 |
| 12 | −.91607978 | 3.7979295 | 2.6940900 | 4.9347939 |
| 13 | −.91607978 | 3.8214086 | 2.6592893 | 4.9464483 |
| 14 | −.91607978 | 3.8215257 | 2.6592462 | 4.9464398 |

TABLE 5.1.3

SOLUTION OF RICATTI EQUATION FOR FIRST EXAMPLE AT "TIME TO GO" = 12 SECONDS COMPARED WITH HALF THE $P$ MATRIX FROM THE ITERATIVE SCHEME

| | | | |
|---|---|---|---|
| $w_{11}$ | 4.7249809 | 4.7249865 | $\frac{1}{2} p_{11}$ |
| $w_{12}$ | 5.1607696 | 5.1607780 | $\frac{1}{2} p_{12}$ |
| $w_{13}$ | 2.2681971 | 2.2681011 | $\frac{1}{2} p_{13}$ |
| $w_{14}$ | 0.0916074 | 0.0916080 | $\frac{1}{2} p_{14}$ |
| $w_{22}$ | 9.3299679 | 9.3299820 | $\frac{1}{2} p_{22}$ |
| $w_{23}$ | 5.3879865 | 5.3879945 | $\frac{1}{2} p_{23}$ |
| $w_{24}$ | 0.7367288 | 0.7367302 | $\frac{1}{2} p_{24}$ |
| $w_{33}$ | 4.8829110 | 4.8829160 | $\frac{1}{2} p_{33}$ |
| $w_{34}$ | 0.8258760 | 0.8258769 | $\frac{1}{2} p_{34}$ |
| $w_{44}$ | 0.3524266 | 0.3524268 | $\frac{1}{2} p_{44}$ |

When the iterations were started from unstable control gain vectors, the iterative scheme converged to unstable solutions. Examples are shown in Table 5.1.2.

The matrix Riccati differential equation (35) was solved for the system; when "time to go" had increased to 12 seconds the solution of the equation, $W$, agreed to better than four figures with the $P$ matrix calculated by the iterative scheme. The results are given in Table 5.1.3.

**5.2** *A nonlinear system.* Nonlinear systems may be written in the form: $(d\underline{x}/dt) = A\underline{x} + B\underline{u}$, where $A$ and $B$ depend on the state variables. For our example we have taken the Van de Pol system:

$$x_1' = x_2 ;$$

$$x_2' = u + \epsilon(1 - x_1^2)\, x_2 - x_1 .$$

When put in the form of Eq. (1)

$$A(t_i) = \begin{bmatrix} 0 & 1 \\ -1 & +\epsilon(1 - x_1^2(t_i)) \end{bmatrix}; \qquad B = I;$$

$$\underline{u} = \begin{bmatrix} 0 \\ u \end{bmatrix}; \qquad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix};$$

where $u = \underline{k}^T \underline{x} = k_1 x_1 + k_2(t)\, x_2$.

The objective function, $J$, is taken as

$$J = \min \int_0^T (\underline{x}^T \underline{x} + \lambda \underline{u}^T \underline{u})\, dt,$$

where $T$ is the time for which the process runs. Two separate cases of this system were investigated. First, for comparison, the system used by Merriam (Ref. [4], p. 261), where he takes $T = 5$ sec; $\epsilon = \lambda = 1$; and an initial condition $\underline{x}_0 = [1, 0]$. Using optimal control Merriam has estimated that the value of the objective function $J$ is 2.8669. The corresponding values of the objective function using controls calculated by the successive approximation method described in this paper are shown in Table 5.2.1. Values are given corresponding to different frequencies of updating the linear model. It is apparent from the results that the updating frequency is not critical and a low enough frequency could be used for the on-line calculation of the required control.

As our second case of the Van de Pol system we took the very nonlinear example $\epsilon = 9$; $\lambda = 0.1$; $\underline{x}_0 = [1, 1]$ for $T = 2$ seconds. Figure 1 shows the variation of the element $1 - x_1^2(t_i)$ in the $A$ matrix and indicates why the objective function $J$ is much larger for updating periods greater than

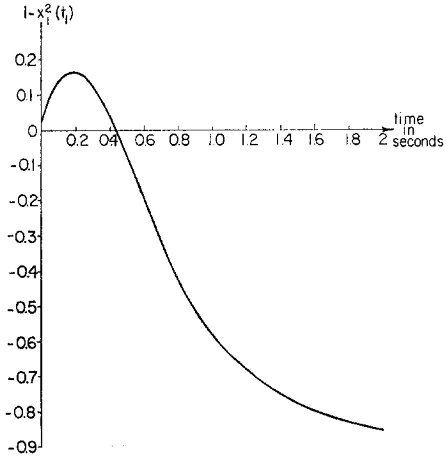


FIG. 1

0.3 sec. The variation in this element only becomes important when $\epsilon$ is large enough for the element to become dominant.

Table 5.2.2 shows that this method of successive approximation provides a good first approximation for this example.

The trajectory for a period of 10 sec was calculated in 44 sec, excluding output time, on an Elliott 503 and the trajectory and control is shown in Fig. 2.

Since the control $u$ is a more rapidly varying function than $k_2$, a greater updating frequency would be required if no state feedback were used.

TABLE 5.2.1

THE EFFECT OF UPDATING FREQUENCY ON THE OBJECTIVE FUNCTION, $J$

| Updating period (sec) | $J$ for $\epsilon = \lambda = 1$ $x_0 = [1, 0]$; $T = 5$ sec | $J$ for $\epsilon = 9$; $\lambda = 0.1$ $x_0 = [1, 1]$; $T = 2$ sec |
|---|---|---|
| 0.01 | 2.9148472 | 3.0245843 |
| 0.05 | 2.9199204 | 3.0943677 |
| 0.10 | 2.9260989 | 3.2036706 |
| 0.20 | 2.9431180 | 3.5337064 |
| 0.30 | 2.9623308 | 4.1117007 |
| 0.40 | 2.9843802 | 39.040771 |
| 0.50 | 3.0088896 | 30.242025 |
| 0.70 | 3.0603105 | 20.353175 |
| 1.00 | 3.1418012 | 29.780394 |

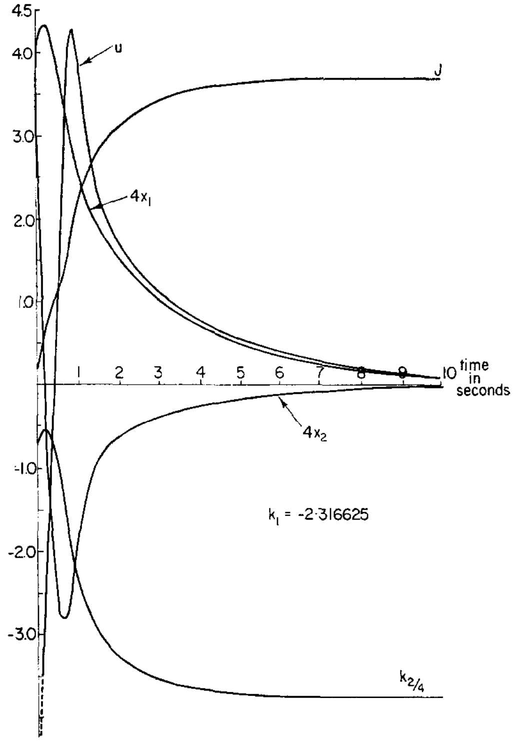


Fig. 2

## 6. Conclusion

The method has been shown to be a fast and accurate way of obtaining the steady state control of linear systems. The example given in 5.1 took

TABLE 5.2.2

Objective Function Values for the Van de Pol System $\lambda = 0.1$. $\epsilon = 9$, $x_0 = [1.1]$. $T = 2$ sec: by Various Methods all Assuming Infinite "Time to Go"

| | |
|---|---|
| Using control given by linear model $x' = A(x = 0)\,x + Bu$; | 29.639976 |
| With no control applied to the system | 20.315419 |
| Using control given by the first two terms of a perturbation series[1] | 3.210048 |
| Using the method described in this paper | 3.024584 |
| Using control given by the first three terms of a perturbation series[1] | 2.557654 |

[1] This work on the pertubation series solution will be published shortly.

5 or 6 iterations at 0.18 sec per iteration to give a result accurate to better than four figures.

The method can be used for on-line control of nonlinear systems by introducing the idea of model feedback in addition to state feedback. The up-dating frequency must be carefully chosen and need not necessarily remain constant for the whole of the trajectory.

Used off-line, the method is capable of providing a first approximation to the optimal control of nonlinear systems with a low computing requirement.

## References

1. R. E. Bellman. "Dynamic Programming." Princeton University Press, Princeton, New Jersey, 1957.
2. R. E. Bellman and R. Kalaba. "Quasilinearization and Nonlinear Boundary-value Problems." American Elsevier, New York, 1965.
3. R. E. Bellman and S. Dreyfus. "Applied Dynamic Programming." Princeton University Press, Princeton, New Jersey, 1962.
4. C. W. Merriam, III. "Optimization Theory and the Design of Feedback Control Systems." McGraw-Hill, New York, 1964.